DOCUMENT RESUME

ED 109 254 TM 004 718

AUTHOR Huberty, Carl J.; Smith, Douglas U.
TITLE Measures of Discrimination Among Achievement Levels in Statistics.
PUB DATE [Apr 75]
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; Classification; *Courses; Grades (Scholastic); *Graduate Students; Graduate Study; *Predictor Variables; Statistical Analysis; *Statistics; Student Characteristics

ABSTRACT

Eight discriminators were identified and data were obtained from the records of 80 graduate students who attained one of four achievement levels at the conclusion of a beginning course in educational statistics. Although the internal discriminatory power of the set of eight measures was very high, estimates of the true power were discouragingly low. Two GRE measures were judged to be the best discriminators, but very poor when considered alone or in combination. Prediction for the second achievement level appeared fairly strong, even for an external analysis. Linear as well as quadratic classification results are included. (Author)

MEASURES OF DISCRIMINATION AMONG ACHIEVEMENT LEVELS IN STATISTICS

Carl J Huberty and Douglas U. Smith

University of Georgia

Paper presented at the Annual Meeting of the American Educational Research Association, Washington, April, 1975.

ABSTRACT

Eight discriminators were identified and data were obtained from the records of 80 graduate students who attained one of four achievement levels at the conclusion of a beginning course in educational statistics. Although the internal discriminatory power of the set of eight measures was very high, estimates of the true power were discouragingly low. Two GRE measures were judged to be the best discriminators, but very poor when considered alone or in combination. Prediction for the second achievement level appeared fairly strong, even for an external analysis. Linear as well as quadratic classification results are included.

Introduction

The academic background of education and psychology graduate students enrolled in beginning statistical methods (or data analysis) courses is sometimes quite varied. In particular, their quantitative skills typically vary from those mastered in beginning high school mathematics to those mastered in the study of calculus. It might be desirable to restrict enrollment in statistical methods courses to those students who have attained a certain mastery level in mathematics. However, statistical methods courses are required of most doctoral students in education and psychology, regardless of their mathematics mastery level. It might also be argued that mastery of mathematics beyond simple algebra is not requisite for the intended understanding to be gained in these courses. Mathematical maturity is but one student characteristic that may contribute to the variability of achievement in graduate level statistical methods courses. Others might be age, past general academic achievement, past specific nonmathematical achievement, and, possibly, personality characteristics. The purpose of this study was to examine those characteristics of graduate students that potentially discriminate among groups of students in various levels of achievement at the conclusion of an introductory course in educational statistical methods.

Method

Subjects

The sample used in this study consisted of graduate students that had completed an introductory course in educational statistics at The University of Georgia offered in the Department of Educational Psychology. Data were collected for classes of students who had enrolled in the course beginning with the Summer Quarter of 1970 and continuing through the Fall Quar-

ter of the 1974-75 academic year. Six classes, with mean size of 13.5 and range of 19-6, were taught by the same instructor (the first author). The content of the course remained fairly stable; approximately the first half was spent on the typical introductory descriptive methods, with the remaining time spent on simple correlation and regression. A total of 81 students was considered in this study. One student (non-degree) was excluded from the study because of incomplete records, reducing the total sample size to 80. As could be determined from the available records, a clear majority (64) of the students had undergraduate training for elementary and/or secondary school teaching. The sample is characterized in more detail in Table 1.

Insert Table 1 about here

As is evident from examining Table 1, this course, the first in a three-course sequence, appears to be primarily a service course for non-Educational Psychology graduate students--this also holds true for classes taught by other instructors. It might be mentioned that some students, particularly those in the fields of statistics and mathematics, start the sequence with the second course.

Variables

Prior to data collection, potential discriminators of student achievement were specified. Files were then examined to determine the information available for each student. Based on specified and available information, thirteen potential discriminators were selected: age of the student (AGE), scores on both the verbal (GREV) and quantitative portions (GREQ) of the Graduate Record Examination, scores on the common (NTEC) and the teaching area (NTET) portions of the National Teacher Examination, the number of hours

of undergraduate level courses in mathematics/statistics (UHMS), the grade point average attained in those courses (UAMS), the number of hours of graduate level mathematics/statistics courses completed prior to the course in educational statistics (GHMS), the grade point average achieved in those courses (GAMS), the number of years since the completion of the last mathematics/statistics course (YCMS), the undergraduate grade point average (UGPA), the total number of graduate hours completed by the student prior to his taking the beginning statistics course (GHRS), and graduate grade point average prior to the course (GGPA).

Since there were only a limited number of students for which four of the measures were available, these measures were excluded from subsequent analyses. The GHMS and GAMS measures were available for only nine of the 80 students; NTEC and NTET measures were available for only 39 and 35 students, respectively. Thus, nine measures remained: AGE, GREV, GREQ, UHMS, UAMS, YCMS, UGPA, GHRS, and GGPA.

One of four levels of end-of-course achievement was recorded for each student: A, B, C, or D. Achievement or grade levels for the course were based on approximately eight quizzes, one test, and a final examination; all three assessment methods were of the multiple-choice variety, and had very nearly the same number of items from class to class. Final course achievement levels were determined by a linear combination of z-scores. Grade level distributions varied somewhat from class to class. For example, in one class approximately 58% was in the A-level and 17% in the C-level, while another class had only 6% in the A-level with 33% in the D-level. The numbers of students in the achievement levels were: A, 17; B, 33; C, 19; and D, 11.

Data Analyses

Preliminary univariate analyses of variance were carried out to identify measures which did not show any promise of contributing ($F<1.00$) to multivariate separation of the four end-of-course achievement level groups. All univariate F values for the nine remaining measures were greater than 1.95; hence all nine measures were retained for the final analyses.

Data records for some students were not complete. Graduate Record Examination scores were not available for 12 students and were estimated. Estimates for the incomplete data were based on the arithmetic mean on each GRE measure for all available scores across all four grade-levels. For 13 students a YCMS measure could not be determined from the records since they had no undergraduate courses in mathematics or statistics. In these cases it was assumed that they had such a course in their senior year of high school. Since these same 13 students had no undergraduate grade point average in mathematics/statistics (UAMS), an additional analysis was carried out using only the 67 students having the UAMS measure.

In the analyses the condition of multivariate normality was assumed to be met; the condition of equality of the four population covariance matrices was assessed using both a chi-square and an F statistic. When appropriate, separation among the four criterion populations in terms of mean vectors was assessed via Wilks' lambda statistic. Values of a distance measure between pairs of centroids were also obtained to verify the A, B, C, D "ordering" of the four grade levels, and to examine the centroid configuration. Such an ordering was used to detect "second-order" misclassifications--where a student was classified into a grade level nonadjacent to his actual level. Also, an attempt was made to sort out the best and

poorest discriminators, in terms of contribution to group separation.

Classification procedures were used to assess the predictive accuracy of the total set and subsets of discriminators. Both "internal" and "external" classification results were considered. Results of an internal classification analysis are those obtained when measures for the students on whom the basic statistics (mean vectors and covariance matrices) were determined are resubstituted to obtain the values for the classification rules. In an external classification analysis statistics based on one set of students are used in classifying "new" students. The external classification method used in this study is an extension of that proposed by Lachenbruch (1967). The procedure for the Lachenbruch method is as follows: Compute the statistics for each of the possible total samples of size 79 obtained by omitting one student's vector of measures from the original total sample of 80, and record for each computation whether the omitted student is misclassified.

The computer program used was one developed by the first author. This program yields linear and quadratic classification results--both internal and external analyses--as well as the usual values of means, covariance matrices, distances, test statistics, and indices for discrimination.

Results

The values of the statistics using p=8 and N=80 are reported in Table 2. The F values are based on all 80 students, using estimated measures

Insert Table 2 about here

where necessary.

Based on values of test statistics obtained, the condition of equality of the four population covariance matrices was judged untenable -- the ob-

served value of a chi-square statistic (df = 108) was 151.40, $p<.01$; the value of an F statistic (df = 108, 5299) was 1.26, $p<.05$. Because of this conclusion, the appropriateness of the interpretation of Wilks' separation index (the value of which was $\Lambda=0.297$) may be questionable. Distances between pairs of groups based on a pooled covariance matrix verified the ordering of the grade levels. The means for the four levels on the single significant linear discriminant function (LDF) were 9.07, 7.94, 7.55, and 6.85, respectively. Distance-like measures ("likelihood distances") based on separate group covariance matrices also supported the ordering. The usual indices of relative predictor variable contribution -- predictor-LDF correlations, or standardized LDF weights -- must be interpreted with caution. In light of the difficulty of interpretation, all indicators -- correlations, weights, univariate F-values -- suggested that GREQ and GREV were the best predictors, and that GHRS and YCMS were the poorest.

The unequal covariance structure suggested that a nonlinear classification rule be employed. Defining

$$D_{ik}^2 = [(\underline{X}_i - \bar{\underline{X}}_k)' \, S_k^{-1} \, (\underline{X}_i - \bar{\underline{X}}_k)]$$

to be the square of the distance from the point in eight-space representing student i ($\underline{X}_i$) to the point representing the means of the eight measures in group k ($\bar{\underline{X}}_k$), where S_k is the sample (8x8) covariance matrix for group k, the following "quadratic" classification statistic was used:

$$P_{ik} = \frac{p_k \,|\, S_k|^{-\frac{1}{2}} \exp(-\frac{1}{2}D_{ik}^2)}{\sum_{k'=1}^{k} p_{k'} \,|\, S_{k'} \,|^{-\frac{1}{2}} \exp(-\frac{1}{2}D_{ik'}^2)} ,$$

where p_k is the prior probability of membership in population k. This latter expression represents the (posterior) probability of student i belonging to population k. A student is classified into that population from which the sample yields the largest value of P_{ik}. The value of p_k used in this study is N_k/N, where N_k is the size of the sample selected from population k, and $N=\sum_k N_k$.

The results of the internal and external quadratic classification analyses are given in Table 3. Internal classification yielded a high proportion of overall correct classifications (0.838), whereas this proportion fell considerably with the external analysis (0.388). (The latter proportion is about what would be expected under chance classification.) The only grade level for which predictive accuracy remained somew[illegible] respectable in the external analysis was the B-level -- a drop from 0.88 to 0.61. Since a linear rule -- where the pooled sample covariance matrix, S, replaces the S_k matrices in the quadratic statistic, P_{ik} -- is typically used in classification analyses, such results are also given. Linear classification (see Table 4) yielded poorer overall internal proportion of correct classifications (0.600),

---

Insert Tables 3 and 4 about here

---

but better overall external proportion (0.500). With the linear rule the smallest difference between internal and external results was for the A-level group, 0.76 to 0.71; the proportion for the B-level only dropped from 0.79 to 0.67. Internal classification by the quadratic rule did not yield a single second-order misclassification; the linear rule yielded seven such misclassifications. External classification by the quadratic and linear rules produced eight and nine second-order misclassifications, respectively.

Even though the GREQ and GREV measures appeared to be the best, internal quadratic classification yielded an overall proportion of only 0.450 for GREQ alone and 0.488 for the two used in combination. External classifications using the two GRE measures alone yielded proportions about what would be expected by chance; when used in combination the proportion was slightly higher than that expected by chance. When the UGPA measure was included with the two GRE measures, overall proportions were 0.612 and 0.500 for the internal and external analyses, respectively. Again, relative respectability in terms of classification accuracy only held for the B-level students.

An analysis involving the 67 students for whom the grade point average attained in undergraduate level courses in mathematics/statistics (UAMS) was considered did not yield drastically different results. The test statistics indicated unequal covariance structure ($p<.01$); the value of Λ was 0.444. Again, GREQ and GREV appeared as the best discriminators, with GHRS and AGE the poorest; the UAMS measure was near the middle of the nine measures in terms of relative importance. Overall internal and external quadratic proportions of correct classifications were 0.925 and 0.433, respectively; the corresponding proportions obtained from the linear rule were .716 and .552.

Discussion

Perhaps the most striking finding was the drop in the proportion o correct classifications from the internal analyses to the external analyses. That this was particularly true for the quadratic rule should not be too surprising, since with eight or nine predictor measures, the number of estimated parameters is large relative to the sample sizes. The drop was not nearly as severe for the linear classification rule. Whereas the internal classification might be expected to overestimate the true proportion of correct classifications, the external analysis yields an underestimation (Mich-

aelis, 1973). Even though the classification accuracy across all four grade levels is somewhat evasive -- somewhere between 0.388 and 0.838 or between 0.433 and 0.925 -- the measures considered in this study might be expected to do fairly well for the higher grade levels. Further, an external analysis might be expected to yield better results if the number of predictor measures is reduced to include only the "better" ones, as was found in this study when three rather than all eight measures were used. This is presumably due to the fewer parameters that need be estimated -- 24 with three predictors versus 144 with eight predictors for a quadratic external analysis.

The results of this study might appear to support the contention that GRE measures are good predictors of achievement in graduate school. However, to make predictions on the basis of these measures, to the exclusion of others, may be quite hazardous. Predicted grade levels based on separate GRE measures tended to be lower for students in the high levels and higher for those in low levels. It ought to be mentioned that would the variability of the GRE measures be not as restricted as is typical for students already enrolled in graduate programs, the measures might appear as better predictors.

The addition of undergraduate grade point average in mathematics/statistics (UAMS) did not appreciably affect the predictive accuracy of the set of discriminators. A second-order misclassification resulted for all four analyses -- internal and external, and linear and quadratic -- with the inclusion of UAMS; a student who was in the A-level was predicted to be in the D-level. The student's A-level performance was attributed to her tremendous effort; her UAMS measure was only 1.00.

As mentioned previously, an internal analysis may be expected to overestimate the proportion of correct classifications; this is particularly

true for quadratic classification, as was found to be the case in this study, since covariance matrices characterizing each sample are used. However, a linear rule in this study performed better in an external analysis.

Lastly, it is of some interest to note the trends in the descriptive data on the four groups of students (sec " "\`. For all measures save one, the trends were those that might b.. ,ected; grade level and age, and grade level and years since last mathematics/statistics course are inversely related, while grade level and GRE measures, grade level and grade point averages, and grade level and number of undergraduate hours in undergraduate mathematics/statistics courses are directly related. The one exception is the trend across the grade levels of the number of graduate hours completed prior to the statistics course (GHRS); it appears that the B-level and particularly the D-level students delay longer in taking the course. It turned out that the GHRS measure contributed very little to the separation between the four groups.

References

Lachenbruch, P. A. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics, 1967, 23, 639-645.

Michaelis, J. Simulation experiments with multiple group linear and quadratic discriminant analysis. In T. Cacoullos (Ed.), Discriminant analysis and applications. New York: Academic Press, 1973. Pp. 225-238.

Table 1

Sample Description

Sex	
Male	48
Female	32

Degree Program	
Ed.D	28
Ed.M.	27
Ed.S.	12
Ph.D.	11
Non-Degree	2

Graduate Major		
Education		78
Science	14	
Educational Psychology	14	
Reading	9	
Social Science	7	
Administration	6	
Special Education	5	
Mathematics	5	
Curriculum	5	
Vocational	5	
Other	8	
Non-Education		2

Table 2

Means, Standard Deviations, Univariate F's, and

Within-Groups Correlations[a] (p=8, N=80)

	A	B	C	D	F	GREV	GREQ	UMSH	YCMS	UGPA	GHRS	GGPA
	(N_1=17)	(N_2=33)	(N_3=19)	(N_4=11)								
AGE	27.53 (4.67)	31.24 (7.44)	31.74 (7.02)	34.73 (6.86)	2.69	32	-01	03	78	-08	28	-03
GREV	568.90 (74.98)	510.60 (89.62)	485.60 (99.43)	466.10 (73.14)	4.01		10	-15	33	15	12	07
GREQ	624.40 (63.81)	545.80 (94.04)	518.20 (56.06)	484.50 (87.07)	8.47			19	-11	-26	10	-04
UMSH	23.29 (16.26)	15.12 (15.56)	13.11 (19.24)	9.09 (6.20)	2.13				-27	-12	-19	-14
YCMS	6.00 (4.18)	9.89 (7.12)	10.89 (7.49)	13.55 (5.75)	3.30					-15	24	04
UGFA	3.31 (0.06)	2.99 (0.04)	2.96 (0.03)	2.75 (0.04)	4.04						-14	15
GHRS	51.71 (38.61)	55.09 (29.21)	51.16 (28.39)	77.82 (35.08)	1.95							14
GGPA	3.79 (0.02)	3.72 (0.03)	3.57 (0.03)	3.45 (0.04)	3.80							

<u>Note</u>. Standard deviations are given in parentheses.

[a]Decimal points are omitted.

Table 3

Frequencies and Proportions

of Classifications

(Quadratic Rule)

Internal

Predicted Grade-Level

		A	B	C	D	Total
	A	15(.88)	2	0	0	17
Actual	B	2	29(.88)	2	0	33
Grade-Level	C	0	6	12(.63)	1	19
	D	0	0	0	11(1.0)	11

Overall proportion of correct classifications = 0.838

External

Predicted Grade-Level

		A	B	C	D	Total
	A	8(.47)	8	1	0	17
Actual	B	4	20(.61)	9	0	33
Grade-Level	C	2	15	1(.05)	1	19
	D	0	5	4	2(.18)	11

Overall proportion of correct classifications = 0.388

Note. Main diagonal entries indicate correct classifications; off-diagonal entries indicate misclassifications.

Table 4

Frequencies and Proportions

of Classifications

(Linear Rule)

Internal

Predicted Grade-Level

		A	B	C	D	Total
	A	13(.76)	3	0	1	17
Actual	B	3	26(.79)	3	1	33
Grade-Level	C	0	14	4(.21)	1	19
	D	0	5	1	5(.45)	11

Overall proportion of correct classifications = 0.600

External

Predicted Grade-Level

		A	B	C	D	Total
	A	12(.71)	4	0	1	17
Actual	B	5	22(.67)	4	2	33
Grade-Level	C	1	14	1(.05)	3	19
	D	0	5	1	5(.45)	11

Overall proportion of correct classifications = 0.500

Note. Main diagonal entries indicate correct classifications; off-diagonal entries indicate misclassifications.